Hurricane Prediction with Python

Minwoo Lee
lemin@cs.colostate.edu
Computer Science Department
Colorado State University
Hurricane Prediction

- Hurricane Katrina
  - “We lost everything. Katrina didn’t care if you are poor or rich; all the houses look the same now.”
    - Mississippi resident Penny Dean (People magazine)
  - Death toll: 1836
  - Cost: over 200 billion dollars

- Manual hurricane prediction is cumbersome.
- This presentation describes the prototype solution to automate the process of hurricane prediction using Python.
Hurricane Detection

Automated Surface Observing Systems

NOAA Hurricane Detection Buoy

Reconnaissance aircrafts

Hurricane Satellite Image

Doppler Radar Image
What’s wrong with hurricane prediction?

- Large amount of data
- Long time to predict a hurricane
- Cumbersome process for human
  - Hurricane path prediction is not accurate enough
- Need automatic process for early detection and prediction
Global Forecast System (GFS)

- Global spectral model
- Aviation weather forecast
- Forecast
  - Wind speed and direction
  - Temperature
  - Humidity
  - Precipitation
  - Out to 192 hour
  - Every 6 hour
Hurricane Prediction Using GFS data
Data Preprocessing

- Low resolution GFS Analysis
  - 2 degrees of longitude and latitude interval
  - 100 hPa ~ 1000 hPa
  - Recorded every 6 hours
  - 8 features

- Preprocessing
  - Combine 4 grid cell corners at 11 heights
  - Concatenate into a vector
Analysis of GFS Data

- Eye observation of data shows the clear distinction between hurricane and non-hurricane locations.
Python for Data Preprocessing and Analysis

- **Numpy**
  - Basic data structure for matrix representation and operation
  - Tools for reading and storing files, linear algebra, and matrix manipulation

- **Matplotlib**
  - Plot for the eye observation of data
  - Helpful for choosing machine learning algorithm
Support Vector Machines

- Popular statistical method for classification, regression, and novelty detection
- Searching for the hyperplane that separates the data into two classes with maximum margin
PyML for Classification

- Machine learning library
  - Focuses on SVM and kernel methods
  - Written by Dr. Asa Ben-Hur in CSU
- For classification training and prediction, convert Numpy array to VectorDataSet in PyML
- Linear kernel is enough for the GFS data
Linear Kernel for GFS data

- With linear kernel, the data is clearly separable
Fast Prototyping with Python

- PyGTK and Glade
  - Glade-3
    - Rapid application development tool
- Basemap
  - Support plotting data over map projections
User Interface

- Training source selection
- Time period for training data
- Sampling options and data conversion
- Options for classifier
- Train/store/load classifier
- Predict hurricane w/ trained classifier
Data Conversion, Training, Testing, and Displaying

Data conversion

gdtool.createMat(self.datapath, self.trackpath, key, key, store=finen, undersample=True, genkeyf=True)

Training SVM

data = pyml.VectorDataSet(filen, labelsColumn=0)
self.clssfr = pyml.SVM()
self.clssfr.train(data)

Predicting hurricane with test data

result = self.clssfr.test(
    ml.VectorDataSet(filen, labelsColumn=0))

Projection of prediction results on Basemap

self.map.plot(x, y, 'ro', markersize=5*self.dispSize)
Results: Test Data

- **Train**
  - 4 days of preprocessed data
  - Entire globe atmospheric data
  - July 1st ~ July 4th, 2008

- **Test**
  - August 29th, 2008
  - Decent classification result (>90% overall accuracy) with linear kernel
  - 0 false positive
  - Most false negatives are neighboring locations of actual hurricanes
Prediction Results

- Red circles: Actual hurricanes from track data
- Blue circles: Predicted hurricane locations
Conclusions

- Python provides efficient packages useful for hurricane prediction prototyping.
- Discovery of possibility of applying linear model can simplify the problem greatly.
  - Usual complex feature extraction is not necessary.
- Online classification of GFS data from cloud is a feasible next step. (PyCloud?)
  - Early detection
  - Incremental training of classifier for accuracy
- Path prediction with regression model can be another path for future research.
Acknowledgements

● Dr. Mark DeMaria
   Cooperative Institute for Research in the Atmosphere (CIRA)

● Dr. Asa Ben-Hur
   Computer Science Dept. Colorado State Univ.